Computational analysis of protein phosphorylation networks from the phosphoproteomic data

Yu Xue[1,*], Zexian Liu[1], Han Cheng[1]

[1]Department of Systems Biology, Huazhong University of Science and Technology, Wuhan, Hubei 430074, China

*Email: xueyu@hust.edu.cn.

Phosphorylation is one of the most essential post-translational modifications of proteins, regulates a variety of cellular signaling pathways, and at least partially determines the biological diversity. Recent progresses in phosphoproteomics have identified more than 100,000 phosphorylation sites. However, how to extract useful information from flood of data is still a great challenge. Previously, we developed a GPS (Group-based Prediction System) algorithm, which can predict kinase-specific phosphorylation sites for 408 human kinases in hierarchy. Also, we further adopted protein-protein interaction information as a major contextual filter to reduce false-positive hits. With this strategy, we developed iGPS to predict 188,288 site-specific kinase-substrate relations (ssKSRs) between 9,247 targets and 1,079 PKs for 44,290 phosphorylation sites from the phosphoproteomic data, whereas the protein phosphorylation networks (PPNs) were modeled in five eukaryotic organisms. Based on the results, we observed that the eukaryotic phospho-regulation is poorly conserved at the site and substrate levels. Furthermore, we conducted a large-scale phosphorylation analysis of human liver and experimentally identified 9,719 p-sites in 2,998 proteins. Using iGPS, we predicted a human liver protein phosphorylation networks containing 12,819 potential ssKSRs among 350 PKs and 962 substrates for 2,633 p-sites. Further statistical analysis and comparison revealed that 127 PKs significantly modify more or fewer p-sites in the liver protein phosphorylation networks against the whole human protein phosphorylation network. More recently, this methodology was successfully used for modeling a phosphorylation network in Arabidopsis, whereas statistical results suggested that such a network is highly correlated with plant stress resistant processes. Taken together, our studies contribute to the understanding of phosphorylation mechanisms at the systemic level, and provide a powerful methodology for the general analysis of in vivo PTMs regulating sub-proteomes.